

(11)Publication number : 07-281690  
(43)Date of publication of application : 27.10.1995

(72)Inventor : IMAI ATSUSHI  
TSUGI TORU  
NAKAMURA AKIRA  
SEIYAMA NOBUMASA  
MIYASAKA EIICHI

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-281690

(43) 公開日 平成7年(1995)10月27日

(51) Int.Cl.<sup>9</sup>

G 1 0 L 3/00

識別記号

庁内整理番号

F I

技術表示箇所

H

審査請求 未請求 請求項の数 4 O L (全 11 頁)

(21) 出願番号 特願平6-67250

(22) 出願日 平成6年(1994)4月5日

特許法第30条第1項適用申請有り 平成5年10月5日、  
社団法人日本音響学会発行の「日本音響学会平成5年度  
秋季研究発表会講演論文集-I-」に発表

(71) 出願人 000004352

日本放送協会

東京都渋谷区神南2丁目2番1号

(72) 発明者 今井 篤

東京都世田谷区砧一丁目10番11号 日本放  
送協会 放送技術研究所内

(72) 発明者 都木 徹

東京都世田谷区砧一丁目10番11号 日本放  
送協会 放送技術研究所内

(72) 発明者 中村 章

東京都世田谷区砧一丁目10番11号 日本放  
送協会 放送技術研究所内

(74) 代理人 弁理士 谷 義一 (外1名)

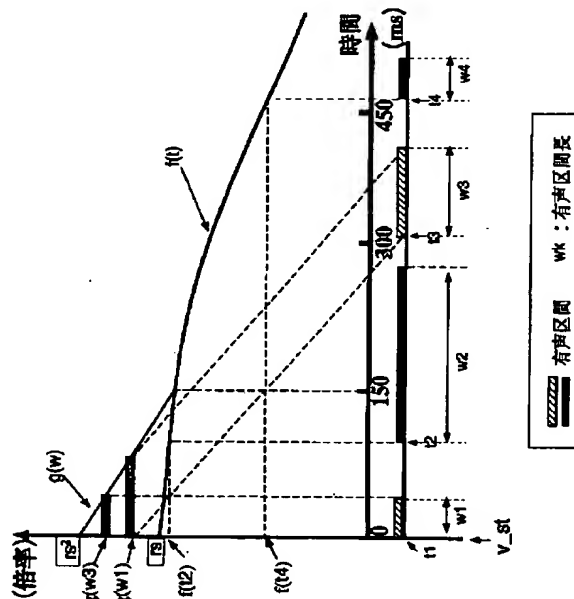
最終頁に続く

(54) 【発明の名称】 話速変換方法

(57) 【要約】

【目的】 各有声区間長の違いにより生ずる聴感上の話速変換効果のばらつきを解消する。

【構成】 有声区間がある値よりも短いものについては聴感上所望の倍率に対応した話速変換効果が得られるように、その区間長に応じてさらに高い倍率で伸張する。例えば、150msに満たないような短い有声音に対しては、図3に示すような倍率関数  $g(w)$  に沿って有声音の長さに対応した伸張倍率を与える、また、一息で発生される区間内で話速を「ゆっくり」から「速く」に変化させるような基準倍率に本発明を適用する場合においては、その区間の開始点から時間450ms以内に出現する上記150ms以下の有声区間について、その出現時刻に関係なく、図3に示すような倍率関数  $g(w)$  に沿って有声音の長さに対応した伸張倍率を与える。150msを越える有声音および経過時間が450msを越える場合は従来の伸張倍率  $f(t)$  を適用する。



## 【特許請求の範囲】

【請求項1】 入力音声の無音区間、無声区間、有声区間を分離し、このうち有声区間を伸張することによって発声の速さ（話速）を声の高さを保ったまま遅くする変換を行う際に、各有声区間の時間長を逐次検出し、各々の有声区間の時間長に一樣な値の、あるいは経過時間とともに滑らかに変化する規準倍率を乗ずることにより、その倍率に対応した聴感的な効果を得る話速変換方法であって、

変換対象となる有声区間の時間長が所定の長さ以上を有する場合にはその有声区間の出現時刻での規準倍率を乗ずるが、変換対象となる有声区間の時間長が、前記所定の長さに満たない有声区間については、その有声区間の時間長に応じて前記規準倍率に比べてより高い伸張倍率を乗ずることを特徴とする話速変換方法。

【請求項2】 前記所定の長さに当る150ms以下の短い有声区間に対しては、その有声区間の出現時刻に関係なく、前記規準倍率に比べてより高い伸張倍率を供する倍率関数に沿ってその有声区間の時間長に対応した伸張倍率を乗じ、また前記150msを越える有声区間の場合は、当該有声区間の時間長に前記規準倍率を乗ずることを特徴とする請求項1に記載の話速変換方法。

【請求項3】 前記経過時間とともに滑らかに変化する基準倍率として、一息で発生する区間を単位にしてこの区間の開始点ではゆっくりとした話速を設定し、その終了に向かって徐々に話速を速めることを特徴とする倍率関数を適用する場合に、この区間の開始時刻から時間450ms以内に出現する前記所定の長さに当たる150ms以下の短い有声区間に対しては、その有声区間の出現時刻に関係なく、前記基準倍率に比べてより高い伸張倍率を供する倍率関数に沿ってその有声区間の時間長に対応した伸張倍率を乗じ、また前記150msを越える有声区間および経過時間が450msを越える場合には、当該有声区間の時間長に前記基準倍率を乗ずることを特徴とする請求項1に記載の話速変換方法。

【請求項4】 前記所定の長さとは規準倍率として実用的な値を設定したときに、変換音声の「ゆっくり感」が聴感的に感じ取れなくなる有声区間の最大時間長を指し、この最大時間長に満たない有声区間については、その時間長 $w$ を変数とする新たな倍率関数 $g(w)$ を導入し、その倍率関数に従って伸張倍率を与えることとし、この倍率関数によって与えられる倍率は前記規準倍率に比べて高い値であって、特に、短い有声区間ほど高倍率になるという性質のものであり、また同倍率関数による倍率の最大増幅値、最低増幅値は固定ではなく、前記規準倍率の規準倍率関数 $f(t)$ の値によってそれぞれが比例的に変化させられるものであることを特徴とする請求項1に記載の話速変換方法。

## 【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、話速変換方法に関し、特に聴覚障害者や高齢者等の音声補聴装置や、一般的な語学学習装置、ラジオ、テープレコーダー、電話などにおいて、話速変換による補助的聴取を行う際の聞き取り易さの向上、テレビジョン、ビデオテープレコーダー、ビデオディスクプレーヤーなどの音声出力を話速変換した際に生ずる映像と音声のズレを効果的に吸収するリアルタイム式の話速変換方法に関する。

## 【0002】

【従来の技術】話速を変換する際に、有声区間を一樣倍率で伸張する手法（中村章ほか平成4年日本音響学会春季研究発表会「高品質リアルタイム話速変換システム」2-6-1P. 329-P. 330（1992-3））や、発声の開始点からの経過時間の関数として倍率を変換にする話速変換法（池沢龍ほか平成4年日本音響学会春季研究発表会「話速変換に伴う時間伸張を吸収するための一手法」2-6-2P. 331-P. 332（1992-3））が存在するが、これらはいずれも各々の有声区間長とは無関係に、伸張倍率を経過時間の関数として一意に定めたものであり、変換音声が全ての有声区間で聴感上同程度の「ゆっくり感」を与えとは言えず、聴感上の効果に「ばらつき」が生ずることがある。従来、これを解決して聴感上安定に且つ自然に変換する話速変換技術はなかった。

## 【0003】

【発明が解決しようとする課題】入力音声の話速を「ゆっくり」にすることを目的として、無音区間、無声区間、有声区間を分離し、無音区間と無声区間の長さはそのままに、有声区間の伸張による話速変換を行う際、音声の中の複数の有声区間を一樣な倍率で伸張した場合、各々の有声区間の区間長によって、聴感上の「ゆっくり感」の程度に差異が生ずることがわかっている（今井篤ほか平成5年日本音響学会秋季研究発表会「話速変換に伴う時間伸張のリアルタイム吸収法」1-9-10P. 361-P. 362（1993-10））。

【0004】音声中には、異なる母音の連鎖や長母音などのように300msを越えるような比較的長い有声区間や、逆に無声区間や無音区間に挟まれた母音などに多い100msを下回るような比較的短い有声区間が相次いで現れることもあり、例えば、この両者が混在する音声に対して、一定の同じ倍率で伸張した音声を取った場合、長い有声区間は1有声区間単位での伸張時間の絶対量が大きく、聴感上の「ゆっくり感」が大きいのに比べ、短い有声区間は伸張時間の絶対量が小さく、場合によっては殆ど「ゆっくり感」が感じられないことがある。

【0005】例えば、区間長が350msと80msのものを従来法により一律に1.5倍に伸張した場合、525msと120msに変換されるが、前者の伸張時間の絶対増加量が175msであるのに対して、後者は僅

か40msの伸張で、これが聴感上の効果の差となって現れてくる。従って、この様に長短さまざまな有声区間分布が一連の入力音声に複数箇所存在する場合は、話速の定まらない不安定な音声に変換されてしまい、場合によってはこれがかなり気になることがある。

【0006】また、既に提案されている、話速変換に伴う時間伸張を吸収する手法(池沢龍ほか平成4年日本音響学会春季研究発表会「話速変換に伴う時間伸張を吸収するための一手法」2-6-2P. 331-P. 332(1992-3))では、一息で発声する区間(フレーズ)の開始点での有声区間の伸張倍率を高く設定し、徐々に話速を速くしていくことで、変換音声の全体としての「ゆっくり感」と、全体としての時間伸張の吸収を実現しているが、このフレーズの開始点付近において短い有声区間が相次いで出現するような音声の場合には、上述の理由により比較的高い倍率を乗じて「ゆっくり感」が得られず、結果的に後半の話速の速い部分だけが目立ってしまうことになり、期待する効果が得られない場合がある。

【0007】上述した問題点を更に具体的事例で示す。20

【0008】(1)一息で発声される区間(フレーズ)の予測長を2000msに固定し、伸張倍率 $r$ を図1に示す曲線に添って $rs$ ( $rs > 1$ )から $re$ ( $re < 1$ )へと単調に減少させる。

【0009】(2)2000msを越えたところではピッチ周波数の変化に伴い倍率に適宜修正を加える。

【0010】この手法をリアルタイム話速変換システムに導入し、多数のニュース音声を変換した結果、いくつかのフレーズについて期待される効果、特に、フレーズの開始点付近において「ゆっくり」した感覚を生じさせる効果の得られないものがあった。図2に、特に効果的であったフレーズ1例(同図の(a))と、特に効果が感じられなかったフレーズ2例(同図の(b)、(c))について、フレーズ内の有声区間長の時間軸上の分布を示す。

【0011】この3例に代表される傾向として以下の点が挙げられる。

【0012】(1)文頭450ms~500ms以内に150msを越える比較的長い有声区間が複数個存在する場合は、伸張倍率 $r$ が $r = 1.4$ でも効果が大きい。40

【0013】(2)フレーズの開始部分に150ms以下の比較的短い有声区間が存在する場合、 $r = 2.0$ でも効果が少ない。

【0014】他のフレーズについても検証した結果、同様の傾向が見られた。

【0015】本発明は、上述した問題点に鑑みてなされたもので、その目的は有声区間の伸張による話速変換を行う際に、入力音声の有声区間長の差異に起因する話速変換効果の聴感上のばらつきを無くし、いかなる入力音声に対しても自然で、且つ安定した話速変換効果が得ら

れる話速変換方法を提供することにある。

【0016】

【課題を解決するための手段】上記目的を達成するために、本発明は、入力音声の無音区間、無声区間、有声区間を分離し、このうち有声区間を伸張することによって発声の速さ(話速)を声の高さを保ったまま遅くする変換を行う際に、各有声区間の時間長を逐次検出し、各々の有声区間の時間長に一樣な値の、あるいは経過時間とともに滑らかに変化する規準倍率を乗ずることにより、その倍率に対応した聴感的な効果を得る話速変換方法であって、変換対象となる有声区間の時間長が所定の長さ以上を有する場合にはその有声区間の出現時刻での規準倍率を乗ずるが、変換対象となる有声区間の時間長が、前記所定の長さに満たない有声区間については、その有声区間の時間長に応じて前記規準倍率に比べてより高い伸張倍率を乗ずることを特徴とする。

【0017】また、本発明は好ましくは、前記所定の長さに対する150ms以下の短い有声区間に対しては、その有声区間の出現時刻に関係なく、前記規準倍率に比べてより高い伸張倍率を供する倍率関数に沿ってその有声区間の時間長に対応した伸張倍率を乗じ、また前記150msを越える有声区間の場合は、当該有声区間の時間長に前記規準倍率を乗ずることを特徴とすることができる。

【0018】また、本発明を適用する前記経過時間とともに滑らかに変化する基準倍率として、一息で発生する区間を単位にして、この区間の開始点ではゆっくりとした話速を設定し、その終了点に向かって徐々に話速を速めることを特徴とする倍率関数を適用する場合に、上記区間の開始時刻から一定時間内、好ましくは時間450ms程度以内に出現する前記所定の長さに当たる区間長、好ましくは150ms程度に満たない区間長を有する有声区間に対しては、その有声区間の出現時刻に関係なく、前記基準倍率に比べてより高い伸張倍率を供する倍率関数に沿ってその有声区間の時間長に対応した伸張倍率を常時、また前記150msを越える有声区間および経過時間が450msを越える場合は、当該有声区間の時間長に前記基準倍率を乗ずることを特徴とすることができる。

【0019】なお、上記の150ms、450msの値は好ましい値の1つを具体的に例示したものであって、本発明はこの値に限定されるものではない。

【0020】また、本発明は好ましくは、前記所定の長さとは規準倍率として実用的な値を設定したときに、変換音声の「ゆっくり感」が聴感的に感じ取れなくなる有声区間の最大時間長を指し、この最大時間長に満たない有声区間については、その時間長 $w$ を変数とする新たな倍率関数 $g(w)$ を導入し、その倍率関数に従って伸張倍率を与えることとし、この倍率関数によって与えられる倍率は前記規準倍率に比べて高い値であって、特に、

短い有声区間ほど高倍率になるという性質のものであり、また同倍率関数による倍率の最大増幅値、最低増幅値は固定ではなく、前記規準倍率の規準倍率関数  $f(t)$  の値によってそれぞれが比例的に変化させられるものであることを特徴とすることができる。

#### 【0021】

【作用】本発明では、話速変換の効果の程度に影響する有声区間の長さに着目し、聴感上自然で、且つ安定な話速変換効果が得られるように、ある一定の長さ以下の短い有声区間に対しては、その区間長に対応して短いものほどより高い伸張倍率を与えるような新たな関数を適用する。これにより、多様な入力音声をもつて話速変換することが可能になる。特に、従来提案されている話速変換による時間伸張を吸収する手法（池沢龍ほか平成4年日本音響学会秋季研究発表会「話速変換における時間伸張吸収のリアルタイム化の検討」2-9-2P. 349-P. 350（1993-10））に適用した場合は、発声の開始点付近の「ゆっくり感」が聴感上不安定であったという欠点が解消され、安定したより効果的な変換音声を得ることが可能となる。

#### 【0022】

【実施例】以下、図面を参照して本発明の実施例を詳細に説明する。

【0023】一息で発声されると予測される区間内で、この区間の開始点に於いては原音声の話速より「ゆっくり」とした話速を設定し、終了点に向かって一定の規則に従って話速を速めていく「リアルタイム話速変換方法」の手法（池沢龍ほか平成4年日本音響学会秋季研究発表会「話速変換における時間伸張吸収のリアルタイム化の検討」2-9-2P. 349-P. 350（1993-10））に本発明を適用した場合の実施例について説明する。この「リアルタイム話速変換方法」は、実時間で動作する話速変換装置において時間伸張を吸収する手法であるが、これは各フレーズの開始点での「ゆっくり感」が特に要求されるものである。これに本発明による手法を適用することは特に効果的であるといえる。

【0024】図3は本発明の一実施例の動作の概要を示す。文頭から450ms以内に出現する150ms以下の短い有声音に対しては、その有声音の出現時刻に関係なく、図3に示すような倍率関数  $g(w)$  に沿って有声音の長さに対応した伸張倍率を与える。150msを越える有声音および経過時間が450msを越える場合は、従来の伸張倍率曲線  $f(t)$ （図1）を適用する。

【0025】図4～図7は本発明の一実施例を更に詳細に示す図である。

【0026】図4は本発明の一実施例の全体回路構成を示すブロック図である。

【0027】図4に示すリアルタイム話速変換装置は、音声入力回路1と、CPU（中央処理ユニット）回路2

と、PROM（プログラマブルROM）回路3と、入力バッファ回路4と、処理バッファ回路5と、ファイル回路6と、音声出力回路7と、バス8とを備えている。そして、音声入力回路1によって話速変換対象となる音声（原音声）を取り込み、リアルタイム処理で、原音声の声の高さ（ピッチ周波数）の変化を検出すると共に、この検出結果に基づいて、声の高さの高い部分では話速を緩め、低い部分では話速を速めるという規則で話速を変化させることにより、原音声の発話時間を保ったまま、原音声を聴き易い良好な音声に変換する。

【0028】音声入力回路1は、原音声を入力するための一般的な構成の回路、例えばマイクロフォン、音調回路、A/D（アナログ/デジタル）変換器、音声記憶再生回路、音声記録媒体（例えば、ICメモリ、ハードディスク、フロッピーディスクまたはVTR（ビデオテープレコーダ））、およびインターフェイス回路等を備えており、話速変換対象となる音声を取り込み、これをデジタル形式の音声信号に変換するとともに、この変換したデジタル音声信号をCPU回路2からの指示に基づいてフレーム単位で入力バッファ回路4に供給する。

【0029】入力バッファ回路4は、必要な容量のRAM（ランダムアクセスメモリ）などによって構成され、CPU回路2の作業域として使用される部分であり、音声入力回路1から出力される音声信号を取り込んでこれを記憶するとともに、CPU回路2からの指示に基づいて記憶している音声信号を処理バッファ回路5に転送する。

【0030】処理バッファ回路5は、必要な容量のRAMなどによって構成され、CPU回路2の作業域として使用される部分であり、入力バッファ回路4から出力される音声信号を取り込んでこれを記憶するとともに、CPU回路2からの指示に基づいて記憶している音声信号をファイル回路6などに転送する。

【0031】ファイル回路6は、RAMのほかに、ICメモリやフロッピーディスク等の音声記録媒体によって構成され、本発明に係わる有声区間の伸張された音声信号と、無音区間の短縮の処理を施された信号などを格納するメモリであって、処理バッファ回路5から処理済の音声信号が出力されたとき、これを取り込んで記憶し、その後CPU回路2からの指示に基づいて記憶している音声信号を音声出力回路7に供給する。

【0032】音声出力回路7は、ファイル回路6内の音声信号を外部に出力するための一般的な構成の回路、例えばインターフェイス回路、D/A（デジタル/アナログ）変換器、スピーカー、録音装置（あるいは放送機器）等を備えており、ファイル回路6から音声信号が出力されたとき、これを取り込んで音声に変換しながら、外部に出力する。

【0033】また、CPU回路2は、ワンチップマイクロコンピュータ等によって構成される部分であり、PR

OM回路3に格納されている図5、図6に示すようなプログラムに基づいて装置全体の制御や各種のデータ処理を行う。

【0034】また、PROM回路3は、CPU回路2の動作を規定するプログラムや各種の処理で使用される定数データなどの格納場所として使用される部分であり、CPU回路2からの読みだし指令に応じて記憶しているプログラムや定数データを読み出してCPU回路2に供給する。

【0035】次に、本発明の一実施例の動作について図5、図6を参照して説明する。

【0036】図5、及び図6は処理の流れを示すフローチャートであり、図6は図5のST9の有声区間処理ルーチンの詳細を示す。

【0037】ここでは、説明のために音声信号中の息継ぎ区間を「ポーズ」、一息で発生される区間を「フレーズ」、また「フレーズ」の時間長の平均的な値を「予測フレーズ長」と呼び、次のように定義する。

【0038】ポーズ：無音部分と判定された区間のうち、その区間長が $T_{h1}$ （本実施例では $T_{h1}=200$ ms）以上の無音区間。なお、 $T_h$ はスレッシュホールド値を意味する。

【0039】フレーズ：ポーズと次のポーズに挟まれる区間。この区間の開始点を $Ph\_st$ とする。

【0040】予測フレーズ長：フレーズの平均的な時間長で、 $T$ （単位はms）とする。（本実施例では $T=2000$ msとした）また、図6中の $f(t)$ と $g(w)$ は有声区間の伸張倍率を定める関数であり、以下の特性を有するものである。

【0041】 $f(t)$ ：話速変換に伴う時間伸張を吸収するために用いる倍率関数であって、予測フレーズ長内の有声区間の出現時刻 $t$ （ $0 \leq t \leq T$ ）に対して倍率を定める単調減少関数である。

【0042】 $t=0$ におけるあらかじめ定めた倍率を $r_s$ 、 $t=T$ におけるあらかじめ定めた倍率を $r_e$ （ $r_s \geq r_e$ ）とすると、 $f(t)$ は $r_s \geq f(t) \geq r_e$ 、 $0 \leq t \leq T$ を満たす。

【0043】 $g(w)$ ：一定の区間長 $W_i$ （本実施例では $W_i=150$ ms）に満たない有声区間を、その区間長 $w$ に応じて、 $f(t)$ により定まる規準倍率より高い倍率で伸張するための倍率関数であって、有声区間長 $w$ （ $0 < w \leq W_i$ ）に対して倍率を定める単調減少関数である。

【0044】ここで、 $g(w)$ の適用条件を満たした有声区間 $[t_i, t_i + W_i]$ （但し、 $w_i < W_i$ ）に対して、 $g(w)$ の定義により、常に $f(t_i) \leq g(w_i)$ の関係が成り立つ。

【0045】次に、図5の処理手順を説明する。なお、

STはステップを意味する。

【0046】（ST0）まず、 $f(t)$ の最高倍率 $r_s$ と最低倍率 $r_e$ を設定する。

【0047】（ST0-1）次に、フレーム番号 $i$ を0にセットする。

【0048】（ST0-2）続いて、上記 $i$ を $i+1$ とインクリメントする。

【0049】（ST1）そして、音声入力回路1が取り込んだ入力音声を、フレームと呼ばれる一定長の部分に分割し、その結果を入力バッファ回路4に格納する処理を行う。

【0050】本実施例ではフレーム幅6.66msのHamming（ハミング）窓を3.3msずつずらしながら切り出して格納する。

【0051】（ST2）入力音声信号を各フレーム毎に、自己相関法や、零クロス法などの方法で処理して有声、無声、無音の判定を行う。人が発声する有声および無声以外の入力音（例えば、低レベルの雑音や背景音等）は原則として無音として識別処理する。

【0052】（ST3） $i$ 番目のフレームについての有声、無声、無音の判定結果（今回の判定結果）と、 $i-1$ 番目のフレームについて有声、無声、無音の判定結果（前回の判定結果）とが同じであるか否かを判別する。両者の判定結果が同じであれば（ST0-2）に戻り、同じでないならば次の（ST4）に移る。但し、 $i=1$ の場合は（ST0-2）に戻る。

【0053】本実施例では、システム全体の処理の遅延時間を最大限短縮するため、有声、無声、無音の各音声区間については各々の区間長全体を一括して処理するのではなく、出来るだけ短い区間に分割（本実施例では、有声区間を150msに分割）して処理した。

【0054】（ST4） $i-1$ フレームまでの、同じ種類（有声、無声或いは無音）の区間と判定されている音声区間を入力バッファ回路4から処理バッファ回路5に転送して格納する。

【0055】（ST5）処理バッファ回路5に格納されている音声区間が、無音か無声か有声か否かを判定する。無音区間の場合は（ST6）へ進み、無声区間の場合は（ST11）へ移り、有声区間の場合は（ST9）へ移る。

【0056】（ST6）当該無音区間がポーズ区間か否かを判断する。ポーズ区間の場合は（ST6-1）へ移り、ポーズ区間でない場合は（ST8）へ飛ぶ。但し、図4のリアルタイム話速変換装置の起動時はポーズ区間であったと判断し、必ず（ST6-1）へ進む。

【0057】（ST6-1）ポーズ区間以降に出現する有声区間の番号を表す変数 $k$ に初期値としての1を代入する。

【0058】（ST7）ポーズの区間長を調べ、その区間長によって適宜、予め設定されているアルゴリズム

（池沢龍ほか「話速変換に伴う時間伸張を吸収するための一方法」1992年音声研究会P. 49-P. 56）によって聴感上違和感ない程度に短縮する。

【0059】本実施例では、862msを越える区間長を有する無音区間を一律にこの862msの値まで短縮することとし（池沢龍ほか平成4年日本音響学会春季研究発表会「話速変換に伴う時間伸張を吸収するための一手法」2-6-2P. 331-P. 332（1992-3））、無音区間862msを経過した時点で更に無音区間が続く場合は、それ以降の無音データを廃棄して次のフレーズの開始点を待つこととする。

【0060】（ST8）処理バッファ回路5内にある処理済の無音区間の信号をファイル回路6に転送させて格納させた後、処理バッファ回路5をクリアする。次に（ST12）へ移る。

【0061】（ST12）音声信号の最後まで処理したか否かを判定する。肯定判定の場合は本（ST9）の処理をルーチン終了し、否定判定の場合は（ST0-2）へ戻る。

【0062】（ST9）（ST5）で有声区間と判定された区間に対して、後述の図6に示す有声区間処理を行う。この区間の処理における時間軸の原点を $V\_st$ と定義する。また、フレーズ内の第 $k$ 有声区間の開始時刻を $t_i$ 、区間長を $w_i$ と記す。

【0063】（ST9-1）上述の変数 $k$ を $k+1$ とインクリメントする。

【0064】（ST10）処理バッファ回路5内にある話速変換済みの音声データをファイル回路6のメモリに格納するとともに、処理バッファ回路5をクリアする。その後、上述の（ST12）へ移る。

【0065】（ST11）（ST5）において処理対象となる区間が無声と判断されれば、この無声区間の音声信号を処理バッファ回路5からファイル回路6に転送して格納した後、処理バッファ回路5をクリアする。その後、上述の（ST12）へ移る。

【0066】次に、図6のST9有声区間処理ルーチンの詳細を説明する。

【0067】（ST14）まず、有声区間のピッチ抽出を行う。

【0068】（ST15）次に、変数 $k$ が $k=1$ か否かを判定する。 $k=1$ の場合、即ちポーズ区間以降に出現する最初の有声区間の場合は（ST15-1）へ移り、そうでない場合は（ST15-2）へ移る。

【0069】（ST15-1）この有声区間の処理における時間軸の原点を示す変数 $V\_st$ に時刻 $t_i$ を代入する。次に（ST16）へ移る。

【0070】（ST15-2）変数 $k$ が3以下か否か、即ち $k$ が2または3であるか否かを判定する。 $k$ が2または3の場合は（ST16）へ移り、 $k$ が4以上の場合は（ST17）へ飛ぶ。

【0071】（ST16）第 $k$ 有声区間の最大ピッチ周波数を $P_i$ と定義する。 $k=1, 2, 3$ の場合には $P_i$ の値を保存する。

【0072】（ST16-1）変数 $k$ が $k=3$ か否かを判定する。 $k=3$ の場合は次の（ST16-2）へ移り、そうでない場合、即ち $k=1, 2$ の場合は（ST17）へ飛ぶ。

【0073】（ST16-2）3つの有声区間 $P_1, P_2, P_3$ のうちの最大値を

$Pitch\_max$

とする。次に（ST17）へ移る。

【0074】（ST17） $t_i$ が、区間 $[V\_st, V\_st+T]$ に含まれているか否かを判定する。含まれていれば（ST17-1）へ移り、そうでなければ（ST12）へ移る。（本実施例では前述のように $T=2000ms$ とした。）

（ST17-1） $V\_st > t_i$ であるか否かを判定する。

【0075】 $V\_st > t_i$ のときは、発声の終了点間近で意味的重要度が低い場合が多いため、本実施例では特に $g(w)$ を適用せず、（ST17-1）から直接（ST19）に移ることとした。それ以外のときは次の（ST18）へ移る。

【0076】（ST18）フレーズの開始部において変換により生じる聴感上の「ゆっくり感」を効果的にするために必要な時間長を $T_i$ とする。 $T_i$ は実験結果から（今井篤ほか 平成5年日本音響学会秋季研究発表会「話速変換に伴う時間伸張のリアルタイム吸収法」1-9-10P. 361-P. 362（1993-10）） $T$ の $1/4$ 程度が望ましく、本実施例では $T_i=450ms$ とした。

【0077】本処理ブロックでは、第 $k$ 有声区間の終了時刻 $t_i + w_i$ が区間 $[V\_st, V\_st+T_i]$ に含まれているか否かを判定する。含まれていれば次の（ST18-1）へ移り、そうでなければ（ST19）へ移る。

【0078】（ST18-1） $k$ 番目の有声区間長 $w_i$ と、予め設定されている区間長 $W_i$ が、 $w_i \leq W_i$

であるか否かを判定する。肯定判定のときは（ST20）へ移り、否定判定のときは（ST19）へ移る。

【0079】有声区間の伸張による話速変換では、区間長が短いもの程その変換効果が小さくなるが、 $W_i$ は、入力音声をもとに3倍程度の一様な倍率で話速変換した際に、聴感上の話速変換効果が余り感じられなくなる臨界有声区間長を実験により導いた値で、本実施例では $W_i=150ms$ とした。

【0080】（ST19）予め設定した倍率関数 $f(t)$ を適用して有声区間を伸張する。この $f(t)$ は単調減少関数であり、本実施例では以下の式（1）のよ

うな余弦関数を用いて、倍率を $r_s$ から $r_e$ まで変化させた。

【0081】(図7の曲線のグラフ参照)

$$f(t) = r_e + 0.5(r_s - r_e) \{ \cos \pi (t - V_{st}) / T + 1.0 \} \quad (1)$$

但し、 $V_{st} \leq t \leq V_{st} + T$

本実施例では、 $1.0 \leq r_s \leq 1.6$ ,  $0.7 \leq r_e < 1.0$ の範囲で任意に値を定めた。その後、図5のメインルーチンに戻る。

【0083】(ST20)  $V_{st}$ からの経過時間にかかわらず、当該有声区間の区間長 $w_i$ に対して、 $g(w_i)$ で定まる倍率を適用して有声区間を伸張する。

$$g(w) = (- (r_s^2 - f(W_i)) w / W_i) + r_s^2 \quad (2)$$

但し、 $V_{st} = 0$ で $g(W_i) = f(W_i)$ とした。

【0087】(ST21) 処理対象となっている有声区間の最大ピッチ周波数 $P_k$ が、以下の式(3)の条件を満たす場合は(ST22)へ、満たさない場合は(ST★

$$P_k > P_{pitch\_max} \times Th2 \quad (3)$$

本実施例では、 $Th2 = 0.7$ とした。

【0089】(ST22) 変数 $V_{st}$ に時刻 $t_i$ を代入する。

【0090】(ST22-1) 変数 $r_s$ に $(r_s - Th3)$ を代入する。

【0091】これによって、 $f(t)$ は $(r_s - Th3)$ から $r_e$ まで倍率を変化させる。本実施例では、 $Th3 = 0.1$ に設定した。その後、上記の(ST17)へ戻る。

【0092】(ST23) 有声区間を伸張倍率を $r_e$ で伸張する。つまり、話速を最も速い状態のままにする。その後、(ST9)の有声区間処理ルーチンを終了し、図5のメインルーチンに戻る。

【0093】

【発明の効果】以上説明したように、本発明によれば、入力音声の無音区間、無声区間、有声区間を分離し、有声区間を伸張することによって発声する速さ(話速)をゆっくりに変換する方法において、全ての有声区間を一定の倍率によって変換した際に、各有声区間長の違いにより生ずる聴感上の話速変換効果のばらつきを解消するため、有声区間がある値よりも短いものについては所望の倍率に対応した聴感上の話速変換効果が得られるように、その区間長に応じて更に高い倍率で伸張するようにしているので、いかなる発声音声に対しても自然で且つ安定した話速変換効果が得られる。即ち、本発明によれば、受聴者の希望にあった話速に安定、且つ自然に変換することが出来る。

【図面の簡単な説明】

【図1】従来法における倍率関数を示すグラフである。

【図2】従来法を適用した場合の1フレーズ内の有声区間長の時間軸上の分布を示すタイミング図である。

【図3】本発明の一実施例の倍率関数を示すグラフである。

\* 【0082】

【数1】

\*

※ 【0084】本実施例で用いた倍率関数 $g(w)$ は次式

(2)に示す一次関数とし、倍率を $g(0)$ から $g(w_i)$ まで変化させた。その後、図5のメインルーチンに戻る。

【0085】(図7の右角の直線のグラフ参照)

【0086】

【数2】

★23)へ移る。

【0088】

【数3】

【図4】本発明によるリアルタイム話速変換方法の一実施例を適用したリアルタイム話速変換装置の回路構成例を示すブロック図である。

【図5】図4に示すリアルタイム話速変換装置の動作例を示すメインフローチャートである。

【図6】図5に示す有声区間処理ルーチンの詳細を示すフローチャートである。

【図7】図4に示すリアルタイム話速変換装置に $f(t)$ ,  $g(w)$ の関数を適用した場合の動作例を示すタイミング図である。

【符号の説明】

1 音声入力回路

2 CPU回路

3 PROM回路

4 入力バッファ回路

5 処理バッファ回路

6 ファイル回路

7 音声出力回路

8 バス

$f(t)$  話速変換に伴う時間伸張を吸収するために用いる倍率関数

$g(w)$  一定の区間長 $W_i$ に満たない有声区間を、その区間長 $w$ に応じて、 $f(t)$ により定まる規準倍率より高い倍率で伸張するための倍率関数

$r_s$  あらかじめ定めた最高倍率

$r_e$  あらかじめ定めた最低倍率

$T$  予測フレーズ長(フレーズの平均的な時間長)

$Ph_{st}$  フレーズ(ポーズと次のポーズに挟まれる区間)の開始点

$V_{st}$  有声区間の処理における時間軸の原点

$P_k$  第 $k$ 有声区間の最大ピッチ周波数

$pitch\_max$  最初の3つの有声区間 $P_1$ ,  $P_2$ ,  $P_3$ のうち最大値

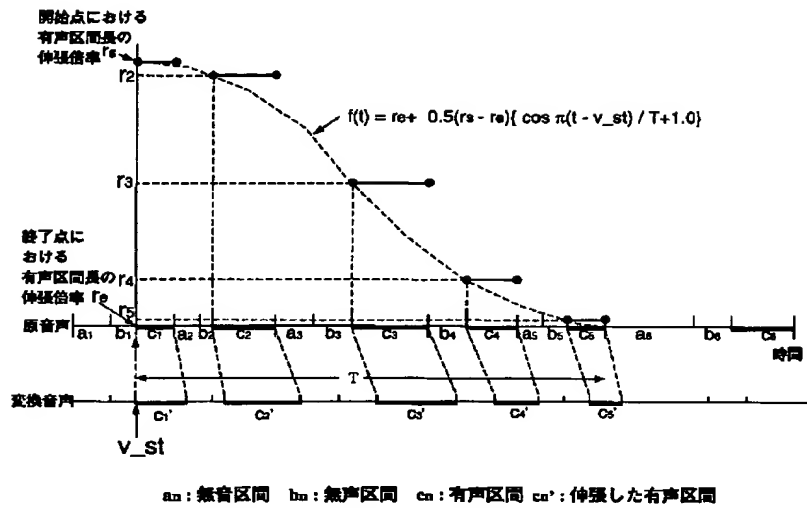


W<sub>i</sub> 予め設定されている区間長  
 w<sub>k</sub> k番目の有声区間長  
 i フレーム番号

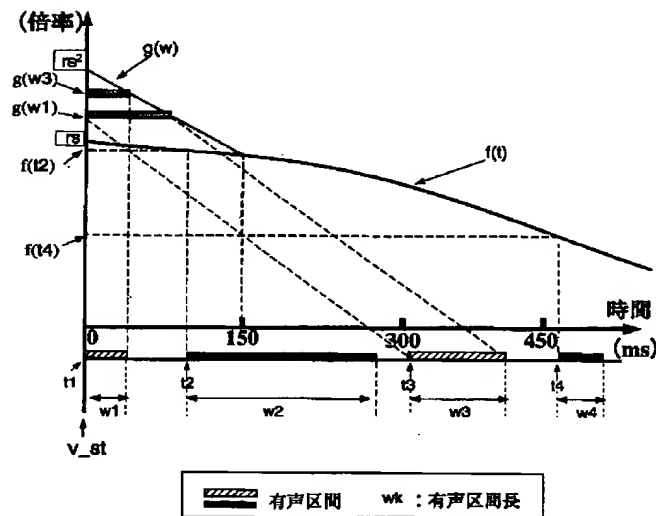
\* k 有声区間番号  
 t<sub>k</sub> 第k有声区間の開始時刻

\*

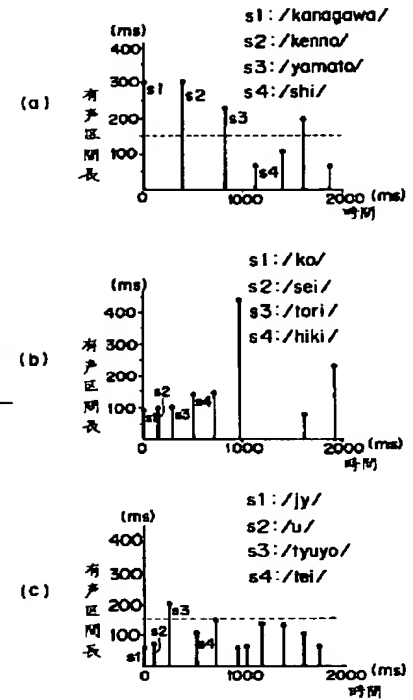
【図1】



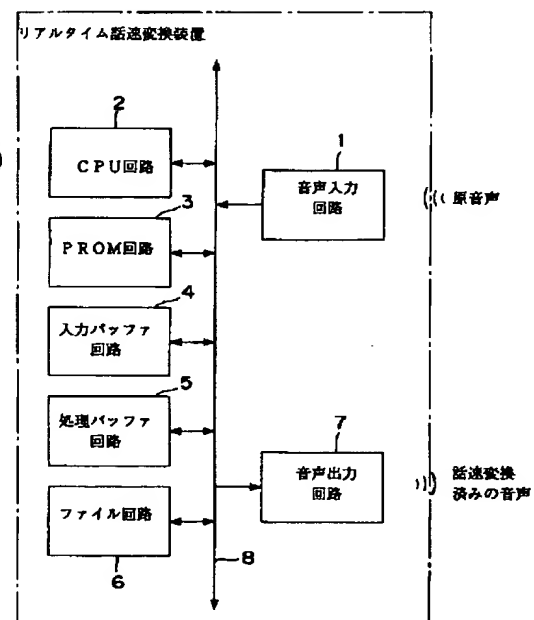
【図3】



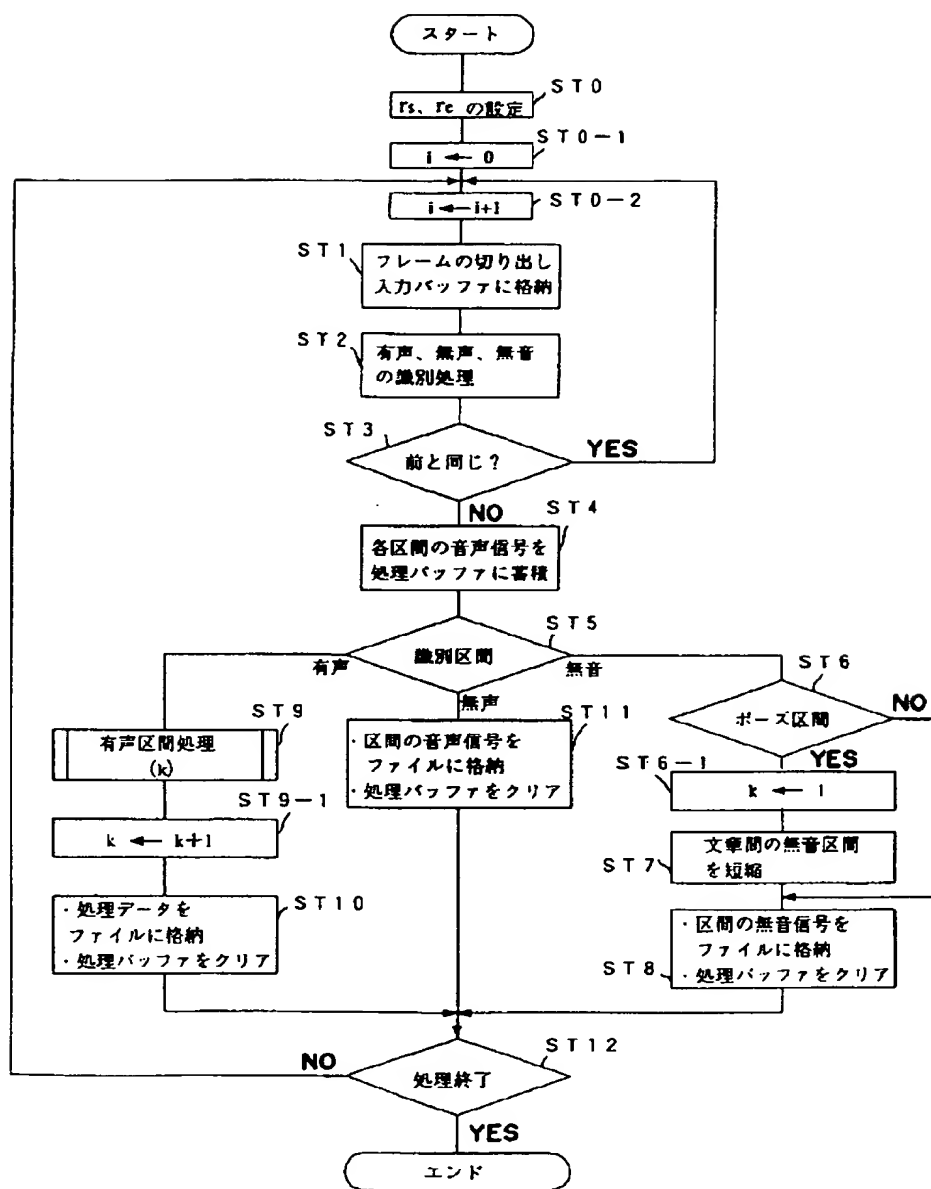
【図2】



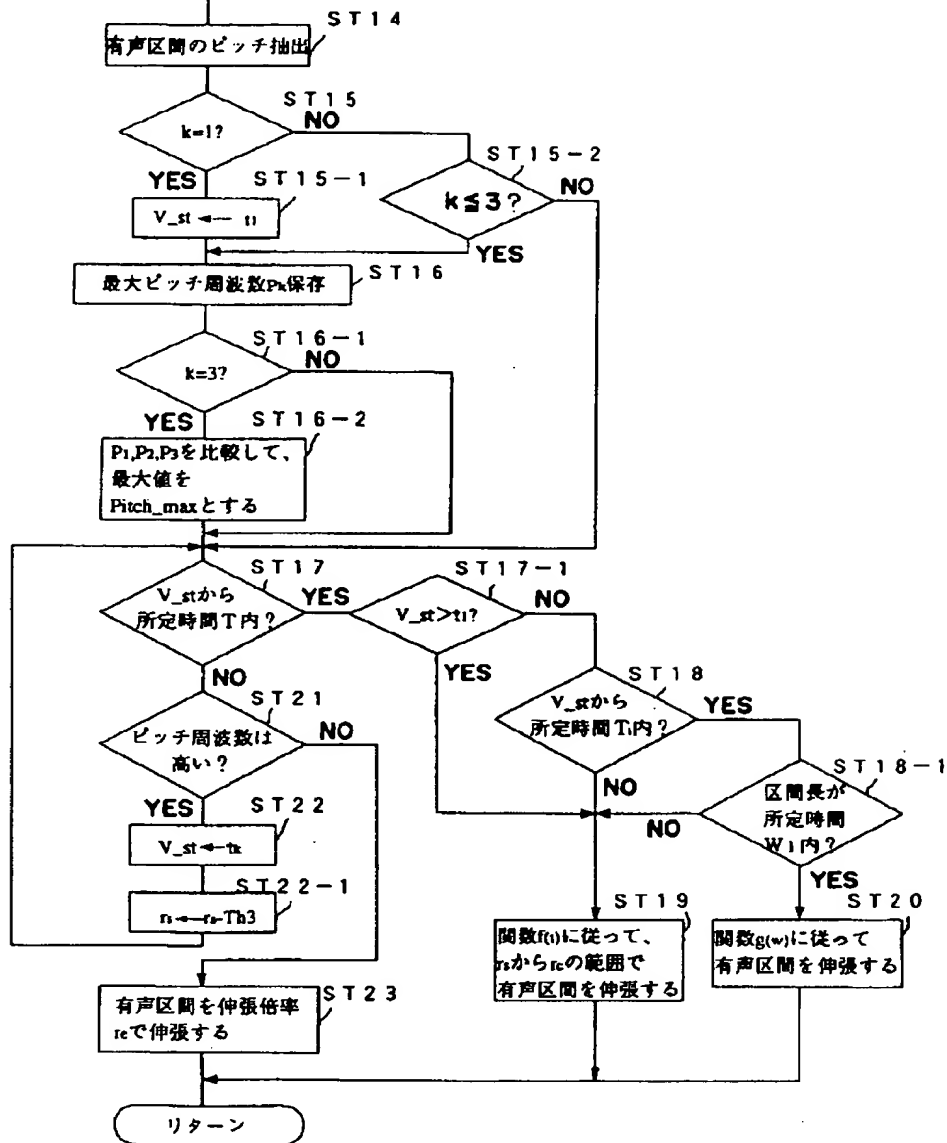
【図4】



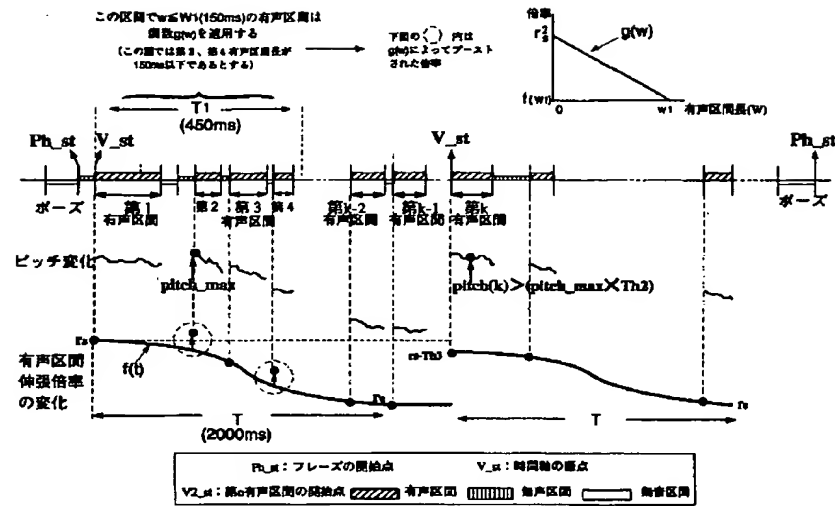
【図5】



**ST9有声区間処理ルーチンスタート**



【図7】



フロントページの続き

(72)発明者 清山 信正  
東京都世田谷区砧一丁目10番11号 日本放  
送協会 放送技術研究所内

(72)発明者 宮坂 栄一  
東京都世田谷区砧一丁目10番11号 日本放  
送協会 放送技術研究所内

【公報種別】特許法第17条の2の規定による補正の掲載  
 【部門区分】第6部門第2区分  
 【発行日】平成11年(1999)8月6日

【公開番号】特開平7-281690  
 【公開日】平成7年(1995)10月27日  
 【年通号数】公開特許公報7-2817  
 【出願番号】特願平6-67250  
 【国際特許分類第6版】  
 G10L 3/00  
 【F1】  
 G10L 3/00 H

【手続補正書】  
 【提出日】平成10年8月25日  
 【手続補正1】  
 【補正対象書類名】明細書  
 【補正対象項目名】発明の名称  
 【補正方法】変更  
 【補正内容】  
 【発明の名称】 話速変換装置  
 【手続補正2】  
 【補正対象書類名】明細書  
 【補正対象項目名】特許請求の範囲  
 【補正方法】変更  
 【補正内容】  
 【特許請求の範囲】

【請求項1】 入力音声の無音区間、無声区間、有声区間を分離し、このうち有声区間を伸張することによって発声の速さ(話速)を声の高さを保ったまま遅くする変換を行う際に、各有声区間の時間長を逐次検出し、各々の有声区間の時間長に一樣な値の、あるいは経過時間とともに滑らかに変化する規準倍率を乗ずることにより、その倍率に対応した聴感的な効果を得る話速変換装置であって、  
変換対象となる有声区間の時間長が所定の長さ以下か否かを判定する判定手段と、

該判定手段の判定結果により、前記変換対象となる有声区間の時間長が所定の長さを越える場合には、その有声区間の出現時刻での規準倍率を乗ずるが、前記変換対象となる有声区間の時間長が、前記所定の長さ以下の有声区間については、その有声区間の時間長に応じて前記規準倍率に比べてより高い伸張倍率を乗ずる演算手段とを有することを特徴とする話速変換装置。

【請求項2】 前記演算手段は、前記所定の長さに当る150ms以下の短い有声区間に対しては、その有声区間の出現時刻に関係なく、前記規準倍率に比べてより高い伸張倍率を供する倍率関数に沿ってその有声区間の時間長に対応した伸張倍率を乗じ、また前記150msを越える有声区間の場合は、当該有声区間の時間長に前記

規準倍率を乗ずることを特徴とする請求項1に記載の話速変換装置。

【請求項3】 前記経過時間とともに滑らかに変化する規準倍率として、一息で発生する区間を単位にしてこの区間の開始点ではゆっくりとした話速を設定し、その終了に向かって徐々に話速を速めることを特徴とする倍率関数を適用する場合に、前記演算手段は、この区間の開始時刻から時間450ms以内に出現する前記所定の長さに当たる150ms以下の短い有声区間に対しては、その有声区間の出現時刻に関係なく、前記規準倍率に比べてより高い伸張倍率を供する倍率関数に沿ってその有声区間の時間長に対応した伸張倍率を乗じ、また前記150msを越える有声区間および経過時間が450msを越える場合には、当該有声区間の時間長に前記規準倍率を乗ずることを特徴とする請求項1に記載の話速変換装置。

【請求項4】 前記所定の長さとは規準倍率として実用的な値を設定したときに、変換音声の「ゆっくり感」が聴感的に感じ取れなくなる有声区間の最大時間長を指し、この最大時間長以下の有声区間については、その時間長wを変数とする新たな倍率関数g(w)を導入し、その倍率関数に従って伸張倍率を与えることとし、この倍率関数によって与えられる倍率は前記規準倍率に比べて高い値であって、特に、短い有声区間ほど高倍率になるという性質のものであり、また同倍率関数による倍率の最大増幅値、最小増幅値は固定ではなく、前記規準倍率の規準倍率関数f(t)の値によってそれぞれが比例的に変化させられるものであることを特徴とする請求項1に記載の話速変換装置。

【手続補正3】

【補正対象書類名】明細書  
 【補正対象項目名】0001  
 【補正方法】変更  
 【補正内容】  
 【0001】

【産業上の利用分野】本発明は、話速変換装置に関し、

特に聴覚障害者や高齢者等の音声補聴装置や、一般的な語学学習装置、ラジオ、テープレコーダー、電話などにおいて、話速変換による補助的聴取を行う際の聞き取り易さの向上、テレビジョン、ビデオテープレコーダー、ビデオディスクプレーヤーなどの音声出力を話速変換した際に生ずる映像と音声のズレを効果的に吸収するリアルタイム式の話速変換装置に関する。

【手続補正 4】

【補正対象書類名】明細書

【補正対象項目名】0015

【補正方法】変更

【補正内容】

【0015】本発明は、上述した問題点に鑑みてなされたもので、その目的は有声区間の伸張による話速変換を行う際に、入力音声の有声区間長の差異に起因する話速変換効果の聴感上のばらつきを無くし、いかなる入力音声に対しても自然で、且つ安定した話速変換効果が得られる話速変換装置を提供することにある。

【手続補正 5】

【補正対象書類名】明細書

【補正対象項目名】0016

【補正方法】変更

【補正内容】

【0016】

【課題を解決するための手段】上記目的を達成するために、本発明は、入力音声の無音区間、無声区間、有声区間を分離し、このうち有声区間を伸張することによって発声の速さ（話速）を声の高さを保ったまま遅くする変換を行う際に、各有声区間の時間長を逐次検出し、各々の有声区間の時間長に一樣な値の、あるいは経過時間とともに滑らかに変化する規準倍率を乗ずることにより、その倍率に対応した聴感的な効果を得る話速変換装置であって、変換対象となる有声区間の時間長が所定の長さ以下か否かを判定する判定手段と、該判定手段の判定結果により、前記変換対象となる有声区間の時間長が前記所定の長さを越える場合にはその有声区間の出現時刻での規準倍率を乗ずるが、前記変換対象となる有声区間の時間長が、前記所定の長さ以下の有声区間については、その有声区間の時間長に応じて前記規準倍率に比べてより高い伸張倍率を乗ずる演算手段とを有することを特徴とする。

【手続補正 6】

【補正対象書類名】明細書

【補正対象項目名】0017

【補正方法】変更

【補正内容】

【0017】また、本発明は好ましくは、前記演算手段は、前記所定の長さに当る 150ms 以下の短い有声区間に対しては、その有声区間の出現時刻に関係なく、前記規準倍率に比べてより高い伸張倍率を供する倍率関数

に沿ってその有声区間の時間長に対応した伸張倍率を乗じ、また前記 150ms を越える有声区間の場合は、当該有声区間の時間長に前記規準倍率を乗ずることを特徴とすることができる。

【手続補正 7】

【補正対象書類名】明細書

【補正対象項目名】0018

【補正方法】変更

【補正内容】

【0018】また、本発明は好ましくは、前記経過時間とともに滑らかに変化する規準倍率として、一息で発生する区間を単位にして、この区間の開始点ではゆっくりとした話速を設定し、その終了点に向かって徐々に話速を速めることを特徴とする倍率関数を適用する場合に、前記演算手段は、上記区間の開始時刻から一定時間内、好ましくは時間 450ms 程度以内に出現する前記所定の長さ

に当たる区間長、好ましくは 150ms 程度に満たない区間長を有する有声区間に対しては、その有声区間の出現時刻に関係なく、前記規準倍率に比べてより高い伸張倍率を供する倍率関数に沿ってその有声区間の時間長に対応した伸張倍率を常時、また前記 150ms を越える有声区間および経過時間が 450ms を越える場合は、当該有声区間の時間長に前記規準倍率を乗ずることを特徴とすることができる。

【手続補正 8】

【補正対象書類名】明細書

【補正対象項目名】0020

【補正方法】変更

【補正内容】

【0020】また、本発明は好ましくは、前記所定の長さとは規準倍率として実用的な値を設定したときに、変換音声の「ゆっくり感」が聴感的に感じ取れなくなる有声区間の最大時間長を指し、この最大時間長以下の有声区間については、その時間長  $w$  を変数とする新たな倍率関数  $g(w)$  を導入し、その倍率関数に従って伸張倍率を与えることとし、この倍率関数によって与えられる倍率は前記規準倍率に比べて高い値であって、特に、短い有声区間ほど高倍率になるという性質のものであり、また同倍率関数による倍率の最大増幅値、最小増幅値は固定ではなく、前記規準倍率の規準倍率関数  $f(t)$  の値によってそれぞれが比例的に変化させられるものであることを特徴とすることができる。

【手続補正 9】

【補正対象書類名】明細書

【補正対象項目名】図 4

【補正方法】変更

【補正内容】

【図 4】本発明の一実施例のリアルタイム話速変換装置の回路構成例を示すブロック図である。